American Journal of Medical Genetics 64:261–265 (1996)

Fragile X "Gray Zone" Alleles: AGG Patterns, Expansion Risks, and Associated Haplotypes

Nan Zhong, Weina Ju, James Pietrofesa, Daowen Wang, Carl Dobkin, and W. Ted Brown
Department of Human Genetics, New York State Institute for Basic Research, Staten Island, New York

The risk for fragile X "gray-zone" alleles to expand appears to depend on the absence of stabilizing AGGs, which interrupt the CGG repeat region. To characterize such alleles better, we analyzed a series of 101 chromosomes with triplet repeat lengths ranging from 35 to 59 for variations in their AGG interspersion patterns. Among these, 11.9% had 3 AGGs, 59.3% had 2, 24.8% had 1, and 4.0% had 0. An inverse relationship between *FMR1* repeat length and the number of interrupting AGGs was observed. Within the range of 35–44 repeats, 98.7% of alleles were found to have a pure CGG repeat length (P_{CGG}) of less than 33. However, among alleles with 45–59 repeats, 50% were found to have 0 or 1 AGG and a P_{CGG} of more than 33. Thus, gray-zone alleles with 45–59 repeats frequently have a long stretch of pure CGGs and thus are more likely to be unstably inherited than alleles with 35–44 repeats. We found length associations of P_{CGG} with 2 flanking microsatellites, DXS548 and FRAXAC1: a $P_{CGG} \leq 20$ was strongly associated with haplotype 20-19, whereas a $P_{CGG} > 20$ was more strongly associated with the haplotype 25-21. This result could reflect a founder effect or a generalized instability of CGGs and microsatellites.

KEY WORDS: fragile X syndrome, genome instability, triplet repeats, mental retardation

INTRODUCTION

The fragile X syndrome is the most common inherited form of mental retardation [Brown and Jenkins, 1992]. The molecular basis for the syndrome usually is an expansion of a repetitive CGG triplet sequence located in the 5′ untranslated region of the fragile X gene, *FMR1* [Fu et al., 1991; Verkerk et al., 1991; Eichler et al., 1993]. In very rare cases, the syndrome is due to either a deletion of part or all of the *FMR1* gene [Gedeon et al., 1992; De Graaff et al., 1995; Qan et al., 1995] or a point mutation within an RNA binding domain [De Boulle et al., 1993; Siomi et al., 1994]. Triplet expansion above a threshold of approximately 200 repeats results in hypermethylation of the *FMR1* promoter region and a lack of gene expression [Pieretti et al., 1991].

The molecular mechanisms underlying the triplet repeat expansion are poorly understood, although several models have been proposed. Two possibilities are that the expansion could be the result of the loss of AGG interruptions within the CGG repeat, which results in an increased rate of DNA slippage, or of an undefined flanking sequence that promotes expansion [Richards and Sutherland, 1992]. A multiallele model, which describes the CGG expansion process, postulates several distinct allele states based on overall length [Morton and Macpherson, 1992]. In this model, the normal (*N*) allele with fewer than approximately 40 repeats, is stable, the intermediate (*S*) allele with approximately 40–60 repeats is prone to conversion at a low rate to the unstable premutation (*Z*) allele with approximately 60–200 repeats, which in turn has a high likelihood of conversion to the full mutation (*L*) allele. The multiallele model has been generalized to 10 alleles [Kolehmainen, 1994] and to *N* alleles by using empirically derived risk figures [Morris et al., 1995]. However, because it is difficult to determine which of the *S* alleles are likely to be unstable by CGG repeat length determination, it has not be practical to apply these models clinically to determine accurately the risk of expansion for intermediate alleles detected within the "gray zone" of approximately 35–59 repeats.

Several studies have shown that the AGG interruptions in the CGG repeat sequence are polymorphic [Eichler et al., 1994; Hirst et al., 1994; Kunst and Warren, 1994; Snow et al., 1994; Zhong et al., 1995]. Previously, we found that 63% of 54 premutation alleles had none and 37% had only one interrupting AGG, whereas there usually was more than one AGG among normal alleles [Zhong et al., 1995]. Our study included 16 gray-zone alleles with 35–52 repeats. They had significantly

Received for publication September 25, 1995; revision received January 10, 1996.

Address reprint requests to W. Ted Brown, Department of Human Genetics, New York State Institute for Basic Research, 1050 Forest Hill Road, Staten Island, NY 10314.

longer 3′ pure CGG repeat lengths and striking variations in associated *DXS548–FRAXAC1* haplotypes versus 117 normal alleles with fewer than 35 repeats [Zhong et al., 1995]. To estimate more exactly the mutation risk for gray-zone alleles, which clinically present important problems for genetic counseling, we have analyzed the AGG interspersion patterns and the associations with flanking microsatellites of a large sample of such alleles. This study provides a more complete basis for gray-zone-allele risk assessment.

MATERIALS AND METHODS

All males with CGG repeat lengths in the range of 35–60, which had been identified in an ongoing fragile X screening program [Brown et al., 1993; present study], were studied. These males were not known to be directly related to any individual with a known diagnosis of fragile X. Included for comparisons were 186 chromosomes previously analyzed [Zhong et al., 1995]. AGG interspersion patterns were analyzed by Mnl I partial digestion based on polymerase chain reaction (PCR) as described elsewhere [Zhong et al., 1995]. Briefly, DNA amplified from primers A [Fu et al., 1991] and 3 [Brown et al., 1993] was phenol/chloroform-extracted once and resuspended in 7 μl dH_2O. To this was added 5 units of Mnl I restriction enzyme (NEB). Partial digestion was then performed in 10-μl reaction volumes for 50–70 min at 37°C under conditions recommended by the supplier. The digested DNA was separated on SequaGel-6 (National Diagnostics) and electroblotted (Enprotech) onto Nytron+ membrane (Schleicherl & Schuell). The membrane was hybridized with an internal oligonucleotide probe (5′AGT GTT TAC ACC CGC AGC GGG CCG GGG GTT) [Zhong et al., 1995]. CGG repeat lengths and microsatellites DXS548 and FRAXAC1 were analyzed as described elsewhere [Brown et al., 1993; Zhong et al., 1994].

RESULTS

AGG Distribution Among Gray-Zone Alleles

A total of 101 gray-zone X chromosomes including 16 previously studied [Zhong et al., 1995] with repeat lengths of 35–59 were analyzed. Among these, 11.9% had 3 AGGs, 59.3% had 2, 24.8% had 1, and 4.0% had 0 (Fig. 1). For purposes of analysis, we divided the sample into 4 groups based on their overall allele repeat numbers: 1 (35–39), 2 (40–44), 3 (45–49), and 4 (50–59). As indicated in Table I, an inverse relation between allele repeat number and number of interrupting AGGs was generally observed. The longer the allele repeat number, the higher the number with 0 or 1 AGG and the lower the number with 2 or 3 AGGs. In group 1, 21.7% had 3

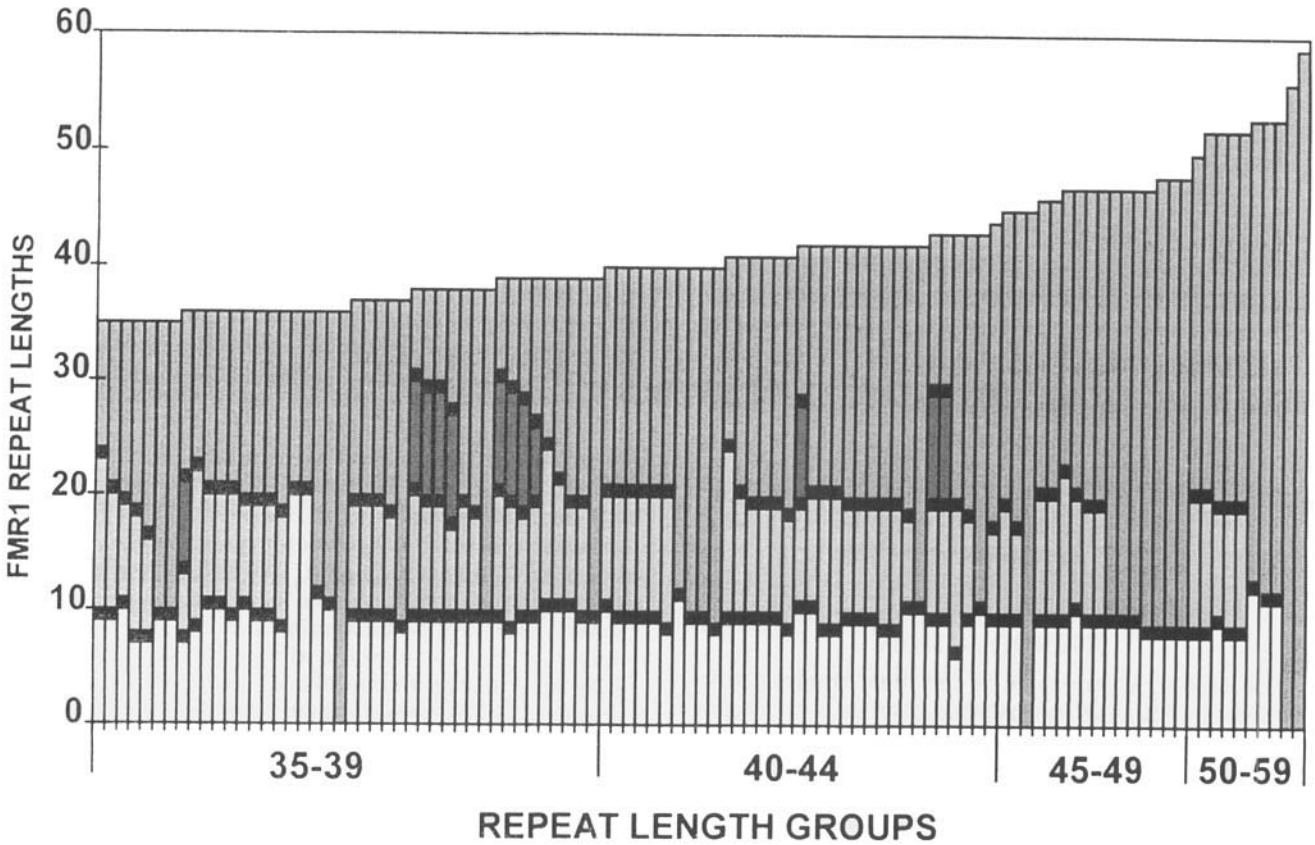

Fig. 1. AGG distribution among 101 gray-zone alleles. Black bars represent the location of the AGGs. The top dark portion of each bar represents the 3′ pure CGG lengths (P_{CGG}s). The repeat length groups correspond to the 4 groups illustrated in Figure 2. Among the alleles in the repeat length range of 45–59, 50% had either 0 or 1 and a P_{CGG} of >33.

TABLE I. AGG Distribution Among Gray-Zone Alleles Grouped by Repeat Lengths

Group	Number	Repeat lengths	% AGGs			
			3	2	1	0
1	42	35–39	21.4	54.8	21.4	2.4
2	33	40–44	9.1	72.7	18.2	0
3	16	45–49	0	50	43.8	6.3
4	10	50–59	0	50	30	20
Total	101	35–59	11.9	59.4	24.8	4.0

AGGs, whereas none within groups 3 and 4 had 3 AGGs. In groups 1 and 2, 1.3% had no AGG, whereas 6.3% of group 3 and 20% of group 4 had no AGG.

We determined the 3′ pure CGG repeat number (P_{CGG}) and divided these into 3 groups: N (normal) with $P_{CGG} < 20$, S (intermediate) with P_{CGG} of 20–32, and U (unstable) with P_{CGG} of 33–59, and compared these P_{CGG} groupings with the number of AGGs and the total repeat numbers. As shown in Table II, increased P_{CGG} values were associated with decreased numbers of AGGs, but, as indicated in Table III, increasing P_{CGG} values were associated with increasing total repeat numbers. Among group 1, N was the most frequent P_{CGG} group (81%). Among group 2, S was the most frequent P_{CGG} group. Among groups 3 and 4, the alleles were equally divided between the S and U groups.

Haplotype Analysis of Gray-Zone Alleles and P_{CGG} Groups

Comparing the allele repeat groupings on the basis of the *DXS548–FRAXAC1* haplotypes (Fig. 2) showed that the fraction with haplotype 25-21 increased as a function of *FMR1* repeat number: 14.3% of group 1, 25.8% of group 2, 33.3% of group 3, and 77.8% of group 4. The uneven distribution of this haplotype was significant (Fisher exact value = 14.0, $P < 0.005$).

Previously, we reported haplotypes of 54 fragile X premutation alleles with repeats of 60–180 [Zhong et al., 1995]. For comparative purposes, this was designated as a separate P_{CGG} group (Z group), with pure CGGs >60. The *DXS548–FRAXAC1* haplotypes found in each of the P_{CGG} groups (N, S, U, Z) are shown in Table IV. Haplotype 20-19 was found in 53.3% of the N group but in fewer than 23% of the other P_{CGG} groups, a significant difference ($\chi^2_{N/S} = 20.4$, $P < 0.001$; $\chi^2_{N/U} = 6.1$, $P < 0.025$; $\chi^2_{N/Z} = 13.7$, $P < 0.001$). Conversely, the haplotype 25-21 occurred significantly less often in the N group (6.7%) than in the other P_{CGG} groups (>20%; $\chi^2_{N/S} = 15.8$, $P < 0.001$; $\chi^2_{N/U} = 5.4$, $P < 0.025$; $\chi^2_{N/Z} = 6.8$, $P < 0.01$).

TABLE II. AGG Distribution Among Gray-Zone Alleles Grouped by Lengths of Pure CGGs (P_{CGG})

Group	P_{CGG}	Number	% AGGs			
			3	2	1	0
N	<20	44	44.3	68.2	4.5	0
S	20–32	43	0	69.8	30.2	0
U	33–59	14	0	0	71.4	28.6

DISCUSSION

Detection of an allele within the gray zone is a common occurrence in screening for fragile X carriers. For example, approximately 1.5% of women randomly selected from a general Caucasian population will have at least one X chromosome with an allele size with ≥50 repeats, 4% with ≥45, 10% with ≥40, and 17% with ≥35 [Brown et al., 1993; 1996]. The present study indicates that most alleles in the range of 35–44 repeats (groups 1 and 2 in Table I) will have one or more AGGs. Such alleles are likely to be stable because they will usually have a P_{CGG} of less than 33 (Fig. 1) [Eichler et al., 1994; Zhong et al., 1995]. Thus, women who carry alleles having fewer than 45 repeats are not likely to be carriers of unstable alleles. However, among the alleles with sizes of 45–59, our results suggest that about half are likely to have no or only one AGG and a P_{CGG} of more than 33 (Table III). Such alleles are potentially unstable. We suggest that alleles with 45 or more repeats are the ones in the gray zone. For technical reasons, we have not been able to analyze the AGG distribution of both X chromosomes from a female. (However, it should be possible to separate the PCR products from a heterozygote and analyze them separately, but overlapping PCR product size prevents the success of this approach.) Therefore, a family study should be performed whenever a woman presents with a gray-zone allele size of 45–59 to determine whether instability exists within the family. The gray-zone alleles analyzed here were not known to be related to any fragile X mutation chromosomes but are being followed up with family stability studies. Only a low level of instability has been seen, but the completed results will be reported separately (S. Nolin et al., personal communication). The empirical risk for a pregnant woman with an allele within this size range to have an affected child appears to be quite small. As far as we know, the smallest allele that has progressed to a full mutation is 58 repeats and the

TABLE III. Relation of P_{CGG} Values and *FMR1* Repeat Groups Among Gray-Zone Alleles Grouped by Repeat Lengths

Group	Number	Repeats	% P_{CGG}		
			N (<20)	S (20–32)	U (33–59)
1	42	35–39	81.0	16.7	2.4
2	33	40–44	30.3	69.7	0
3	16	45–49	0	50	50
4	10	50–59	0	50	50

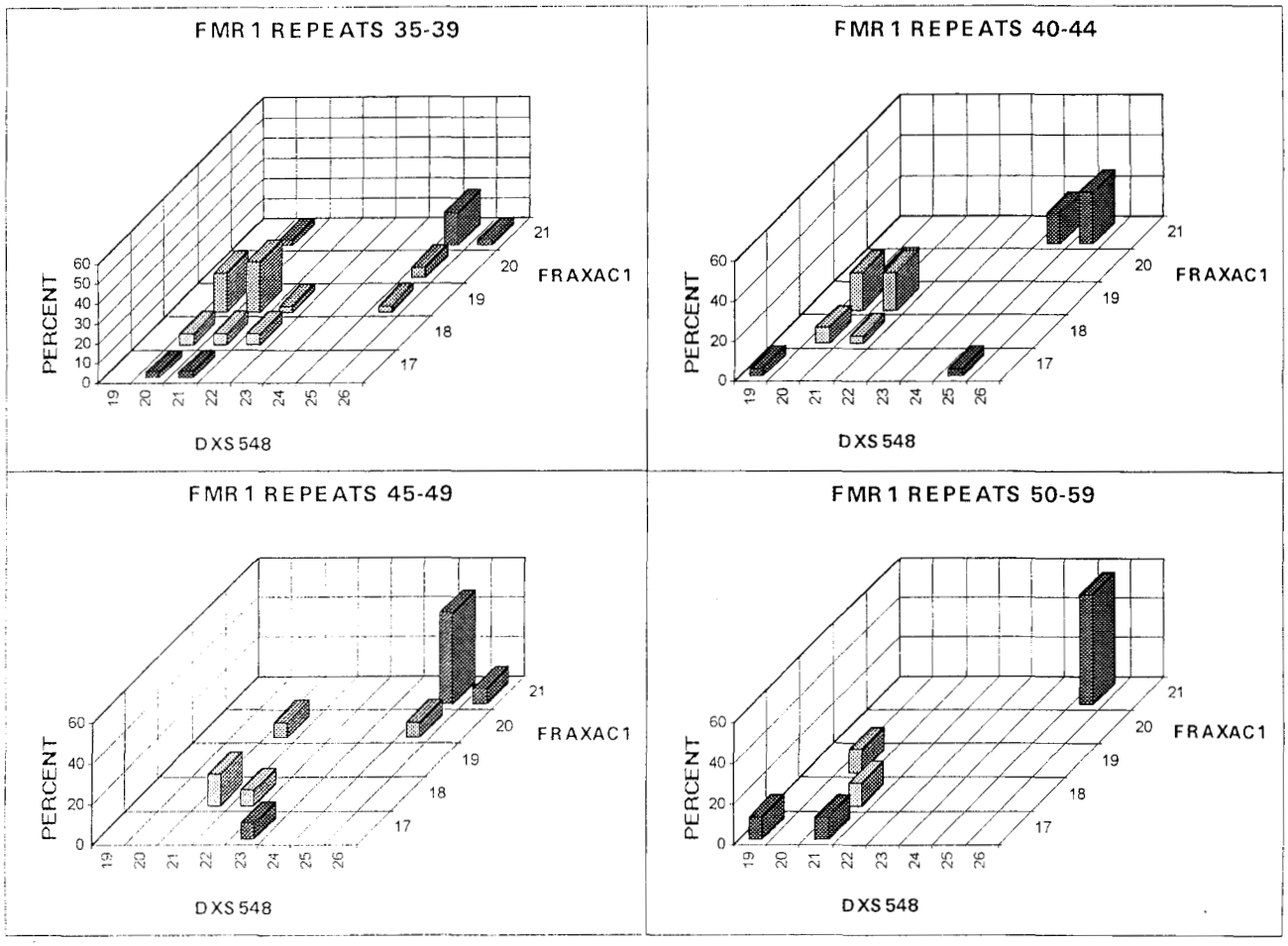

Fig. 2. Association of the gray-zone alleles, grouped by repeat numbers, to *DXS548–FRAXAC1* haplotypes. The 4 groups are separated into 4 different squares. The percentage of the haplotype 25-21 is increased from group 1 to group 4.

smallest size is 55 repeats (J. Holden, personal communication; present study).

By typing the AGG distributions for this series of gray-zone alleles, we were able to analyze the relation of repeat length and P_{CGG} to the *DXS548–FRAXAC1* microsatellite haplotypes. As indicated in Fig. 2 and Table IV, increased allele length and P_{CGG} size are associated with the larger microsatellite alleles. These findings are in agreement with our previous studies, which included a more limited sample of gray-zone alleles [Zhong et al., 1994, 1995]. *DXS548–FRAXAC1* haplotype 20-19 was significantly more common and 25-21 less common for $P_{CGG} < 20$ than for $P_{CGG} > 20$. Similar results were noted by Hirst et al. [1993]: the C7 haplotype (corresponding to our 20-19) was said to "protect" the CGG stability, and the A2 haplotype (25-21) was 7.8 times more common among fragile X chromosomes than among controls.

The risk of having a gray-zone allele that might be unstable and eventually progress to the full fragile X mutation appears to be determined by AGG distribution and P_{CGG}. However, the prior risk of a chromosome that caries the haplotype of 21-18 would appear to be increased, as indicated by Table IV. Eichler et al. [1994] reported that the threshold for stability in their selected series was 34–37 pure CGG repeats. Whether chromosomes with P_{CGG} repeats of this size that also have the 25-21 haplotype are at an increased risk of expansion is not yet known. The present study and previous reports suggest that the fragile X allele is usually associated with larger flanking microsatellites alleles [Richards et al., 1992; Zhong et al., 1993, 1994, 1995; Hirst et al., 1994; Kunst and Warren, 1994; Snow et al., 1994]. We noted that a positive correlation of nearby microsatellite marker repeat sizes exists with each other and with *FMR1* repeat lengths [Brown et al., 1996]. Moreover, we reported previously that repeat instability of a complex microsatellite, FRAXAC2, located ~12 kb downstream from the *FMR1* CGG repeat, appeared to be associated with the fragile X mutation [Zhong et al., 1993]. Instabilities of dinucleotide CA repeats have been found in several inherited cancer syndromes involving DNA mismatch repair genes [Karp and Broder, 1995]. A transgenic mouse model made deficient in a mismatch repair gene (MSH2) [Reitmair et al., 1995] showed susceptibility to lymphomas that

TABLE IV. *DXS548–FRAXAC1* Haplotypes and Numbers (%) of the P_{CGG} Groups

Haplotypes	P_{CGG} groups			
	$N_{<20}$†	S_{20-32}	U_{36-59}	Z_{60-180}†
19–17		1 (1.9)	1 (7.1)	
19–19	2 (1.3)			1 (1.9)
20–17	1 (0.7)			
20–18	15 (10.0)	5 (9.6)		6 (11.1)
20–19*	80 (53.3)	8 (15.4)	2 (14.3)	12 (22.2)
20–20		2 (3.8)		
20–21		1 (1.9)		7 (13.0)
21–17	2 (1.3)		1 (7.1)	
21–18**	9 (6.0)	3 (5.8)	2 (14.3)	8 (14.8)
21–19	8 (5.3)	7 (13.5)		3 (5.6)
21–20		1 (1.9)	1 (7.1)	
21–21				1 (1.9)
22–18	3 (2.0)	1 (1.9)		
22–19	3 (2.0)			
23–17		1 (1.9)		
24–18				1 (1.9)
24–19	3 (2.0)			
24–21	1 (0.7)	5 (9.6)		3 (5.6)
25–17		1 (1.9)		
25–18	2 (1.3)			
25–19	3 (2.0)			
25–20	5 (3.3)		2 (14.3)	1 (1.9)
25–21***	10 (6.7)	15 (28.8)	5 (35.7)	11 (20.4)
26–20	1 (0.7)	1 (1.9)		
26–21	2 (1.3)		1 (7.1)	
Total	n = 150	n = 52	n = 14	n = 54

† Includes data from Zhong et al. [1995].
* $\chi^2_{N/S} = 20.4, P < 0.001$; $\chi^2_{N/U} = 6.1, P < 0.025$; $\chi^2_{N/Z} = 13.7, P < 0.001$.
** $\chi^2_{NS/UZ} = 4.2, P < 0.05$; $\chi^2_{N/U} = 4.9, P < 0.05$.
*** $\chi^2_{N/S} = 15.8, P < 0.001$; $\chi^2_{N/U} = 5.4, P < 0.025$; $\chi^2_{N/Z} = 6.8, P < 0.01$.

were associated with dinucleotide CA repeat instabilities. These studies suggest that DNA mismatch repair genes might play an important role in microsatellite and *FMR1* expansions on fragile X chromosomes. Such DNA mismatch repair gene mutations could be the genetic factor involved in the generation of fragile X mutations, explain the associations of the larger microsatellites alleles with larger gray-zone and fragile X premutations, and account for the founder effects associated with fragile X chromosomes.

ACKNOWLEDGMENTS

This study was supported by grant HD 29407 from the NIH.

REFERENCES

Brown WT, Jenkins EC (1992): The fragile X syndrome. In Friedmann T (ed): "Molecular Genetic Medicine, Vol. 2." San Diego: Academic Press, pp 39–66.

Brown WT, Houck GE, Jeziorowska A, Ding X, Dobkin C, Levinson F, Zhong N, Sklower-Brooks S, Jenkins EC (1993): Rapid fragile X carrier screening and prenatal diagnosis by a non-radioactive PCR test. JAMA 270:1569–1575.

Brown WT, Zhong N, Dobkin (1996): Positive fragile X microsatellite associations point to a common mechanism of dynamic mutation evolution [Letter to the Editor]. Am J Hum Genet 58:641–643.

Brown WT, Nolin S, Houck G, Ding X, Glicksman A, Li S-Y, Stark-Houck S, Brophy P, Duncan C, Dobkin C, Jenkins E (1996): Prenatal diagnosis and carrier screening by PCR. Am J Med Genet 64:191–195

De Boulle K, Verkerk AJMH, Reyniers E, et al. (1993): A point mutation in the *FMR-1* gene associated with fragile X mental retardation. Nature Genet 3:31–35.

De Graaff E, Rouillard P, Willems PJ, Smits APT, Rousseau F, Oostra BA (1995): Hotspot for deletions in the CGG repeat region of *FMR1* in fragile X patients. Hum Mol Genet 4:45–49.

Eichler EE, Richards S, Gibbs RA, Nelson DL (1993): Fine structure of the human *FMR1* gene. Hum Mol Genet 2:1147–1153.

Eichler EE, Holden J, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, Ward PA, Nelson DL (1994): Length of uninterrupted CGG repeats determines instability in the *FMR1* gene. Nature Genet 8:89–94.

Fu YH, Kohl DPA, Pizzuti M, Sutcliffe JS, Richards S, Verkerk AJMH, Holden JJA, Fenwich RG, Warren ST, Oostra BA, Nelson DL, Caskey CT (1991): Variation of the CGG repeats at the fragile X site results in genetics instability: Resolution of the Sherman paradox. Cell 67:1047–1058.

Gedeon AK, Baker E, Robinson H, Partington MW, Gross B, Manca A, Korn B, Poustka A, Yu S, Sutherland GR, Mulley JC (1992): Fragile X syndrome without CGG amplification has an *FMR1* deletion. Nature Genet 1:341–344.

Hirst MC, Knight S, Christodoulou Z, Grewal P, Fryns JP, Davis KE (1993): Origins of the fragile X syndrome mutation. J Med Genet 30:647–650.

Hirst MC, Grewal PK, Davis KE (1994): Precursor arrays for triplet repeat expansion at the fragile X locus. Hum Mol Genet 3: 1553–1560.

Karp JE, Broder S (1995): Molecular foundations of cancer: New targets for intervention. Nature Med 1:309–320.

Kolehmainen K (1994): Population genetics of fragile X: A multiple allele model with variable risk of CGG repeat expansion. Am J Med Genet 51:428–435.

Kunst CB, Warren ST (1994): Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. Cell 77: 853–861.

Morris A, Morton NE, Collins A, Macpherson J, Nelson D, Sherman S (1995): An n-allele model for progressive amplification in the *FMR1* locus. Proc Natl Acad Sci USA 92:4833–4837.

Morton NE, Macpherson JN (1992): Population genetics of the fragile X syndrome: Multiallelic model for the *FMR1* locus. Proc Natl Acad Sci USA 89:4215–4217.

Pieretti M, Zhang F, Fu YH, Warren ST, Oostra BA, Caskey CT, Nelson DL (1991): Absence of expression of the FMR-1 gene in fragile X syndrome. Cell 66:1–20.

Qan F, Zonathan J, Gunter K, Peterson KL, Magenis RE, Popovich W (1995): An atypical case of fragile X syndrome caused by a deletion that includes the *FMR1* gene. Am J Hum Genet 56:1042–1051.

Reitmair AH, Schmits R, Ewel A, Bapat B, Redston M, Mitri A, Waterhouse P, Mittrucker H-W, Wakeham A, Liu B, Thomason A, Griesser H, Gallinger S, Ballhausen WG, Fishel R, Mak TW (1995): MSH2 deficient mice are viable and susceptible to lymphoid tumors. Nature Genet 11:64–70.

Richards RI, Sutherland G (1992): Dynamic mutations: A new class of mutations causing human disease. Cell 70:709–712.

Richards RI, Holman K, Friend K, Kremer E, Hillen D, Staples A, Brown WT, Goonewardena P, Tarleton J, Schwartz C, Sutherland GR (1992): Evidence of founder chromosomes in fragile X syndrome. Nature Genet 1:257–260.

Siomi H, Choi M, Siomi MC, Nussbaum RL, Dreyfuss G (1994): Essential role for KH domains in RNA binding: Impaired RNA binding by a mutation in the KH domain of *FMR1* that causes fragile X syndrome. Cell 77:33–39.

Snow K, Tester DJ, Krucheberg KE, Schaid DJ, Thibodeau SN (1994): Sequence analysis of the fragile X trinucleotide repeat: Implications for the origin of the fragile X mutation. Hum Mol Genet 3: 1543–1551.

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu YH, Kohl DPA, Pizzuti A, Reiner O, et al. (1991): Identification of a gene (*FMR1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65:905–914.

Zhong N, Dobkin C, Brown WT (1993): A complex mutable polymorphism located within the fragile X gene. Nature Genet 5:248–253.

Zhong N, Ye L, Dobkin C, Brown WT (1994): Fragile X founder chromosome effects: Linkage disequilibrum or microsatellite heterozygosity? Am J Med Genet 51:405–411.

Zhong N, Yang W, Dobkin C, Brown WT (1995): Fragile X gene instability: Anchoring AGGs and linked microsatellites. Am J Hum Genet 57:351–361.